IN THE SPECIFICATION

Please amend the Brief Description of the Figures beginning at page 14, line 1 and ending at page 15, line 2 with the following rewritten paragraph:

B1

Figure 14A. Two human sequences with the closest homology to the *C. elegans* sequence gi/1132541 (SEQ ID NO:5, SEQ ID NO:6).

Figure 14B. Computed gene tree indicating that the identified human gene represents an ortholog of the *C. elegans* gene gi/1132541.

Figure 14C. Nucleotide sequence of the death domain gene (SEQ ID NO:7).

Figure 14D. Deduced amino acid sequence of the death domain protein (SEQ ID NO:8).

Figure 15. Identification of candidate gene implicated in the etiology of Chronic Lymphocytic Leukemia (CLL). Sequence homology between a CLL region open reading frame and mouse Rpt1 (sp/P15533/RPT1) is presented (SEQ ID NO:9 and SEQ ID NO.10).

Figure 16A-B. Model of regulatory functions of Rpt1. Figure 16A indicates that in mouse T lymphocytes Rpt1 serves as a repressor of the gene for interleukin 2 receptor (IL-2R). Figure 16B demonstrates that when Rpt1 is knocked out, the regulatory effect is manifested as a block of the apoptotic pathway for T-lymphocytes resulting in accumulation of T-lymphocytes in blood.

Figure 17A. Two EST sequences identified by searching a protein dbEST using the mouse Mad3 protein as a query (SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14).

Figure 17B. Nucleotide sequence of the human Mad3 gene (SEQ ID NO. 15).

Figure 17C. Complete sequence of the human Mad3 protein (SEQ ID NO. 16). A search was conducted to identify overlapping sequences. The complete sequence of the gene was assembled and the amino acid sequence deduced. The translated human Mad3 sequence consists of 206 amino acid residues 81% of which are identical to the mouse Mad3 protein.

B1 Figure 17D. Multiple alignment of the human Mad3 amino acid sequence with known Mad proteins (SEQ ID NOS: 17-22).

Please amend the paragraph beginning at page 27, line 4 and ending at line 13 with the following rewritten paragraph:

B2 This list of gene and protein names is translated into a different alphabet system by substituting each character in the name with a predetermined unique nucleotide combination. The conversion chart is listed in Appendix E. The encoded names are then imported into the BLAST database using the FASTA format. For example, the first entry in the list above is "gfap gamma." After translation using the conversion chart, the same name appears as follows:

AGCAACTAAACACCCATCCAAGCAAACACACACACAAAC (SEQ ID NO: 1)

Thus, the complete FASTA entry looks like this:

E2 cont

>gi|1 species,gp,gfap gamma

AAGCAACTAAACACCCATCCAAGCAAACACACACACAAAC (SEQ ID NO:2)

Please amend the paragraph beginning at page 28, line 8 and ending at line 16 with the following rewritten paragraph:

B3

Thus, the scientific journals are translated, using the same nucleotide combinations, into a continuous string of nucleotides. For example, the sentence "In the absence of costimulation, T cells activated through their antigen ..." is translated into "AAGTACAGATCCACGGAAGGAACGATCCAAACAAAGACGCAACGACAGAAATAAC GATCCACATAACTATCCAAATACATACGCACGGAAGTACACACGTAATTAAACACG GAAGTACATACAGATCCATCCACGGATCCAAATAACGAATTAATTACGCATCCAAA CAAATACGGAAGTACTCAAACACGGAACGAACCATCCACGGAAGGACCTACATACG TAAGCAAGGATCCACGGAAGGAACGAAGTACCTATCCAAACACAGACGGAAGTAA GCAACGACAGATCC " (SEQ ID NO:3).

Please amend the paragraph beginning at page 32, line 13 and ending at page 33, line 2 with the following rewritten paragraph:

B4

Working with nucleotides implies that errors involving reading frames must be addressed. For example, working with a code of four letters, the nucleotide combination ATCTGTCACG (SEQ ID NO:4) could mean ATCT/GTCA or TCTG/TCAC or CTGT/CACG . Since the text is translated into a nucleotide combination, only one of these possibilities is correct. But BLAST can not distinguish between these solutions, *i.e.*, BLAST would potentially

match a database sequence to a wrong reading frame in the query sequence, producing many nonsense results that could compromise the significance of true results.

1. Response to Restriction Requirement

The Examiner has imposed a Restriction Requirement, requesting that Applicants select one of the following groups of claims to pursue in this patent application:

Group I: Claims 1-5 and 7-10, drawn to methods of identifying novel nucleic acid molecules;

Group II: Claim 6, drawn to a method of identifying the effect of a gene knockout on a regulatory pathway;

Group III: Claims 11-21, drawn to methods for extracting information on interactions between biological entities from natural-language text data; and

Group IV: Claims 22-32, drawn to a computer system for extracting information on biological entities from natural-language text data.

In response, Applicants elect to pursue the claims of Group III in this application without prejudice to the prosecution of the subject matter of non-elected claims in other patent applications. Applicants make their election with traverse, on the grounds that the claims of Group III and Group IV are conceptually linked, and would not require separate searches. Accordingly, Applicants request that the restriction requirement be reconsidered.

2. Sequence Listing

Applicants submit herewith a Sequence Listing in computer and paper form, in accordance with 37 C.F.R. §1.821-1.825. The content of the paper and computer readable copies of the Sequence Listing submitted in accordance with 37 C.F.R. §1.821(c) and (e) are the same and do not include new matter.

Rewritten paragraphs including sequence identifiers appear in the preceding "IN THE SPECIFICATION" section. Attached hereto is a marked-up version of the changes made to the specification paragraphs by the instant amendment. The attached page is captioned "VERSION WITH MARKINGS TO SHOW CHANGES MADE" and is only included for the Examiner's convenience. Should any discrepancies be discovered, the version presented in the preceding "IN THE SPECIFICATION" section shall take precedence.

3. Microfiche Copy of Appendix H

On April 14, 2000 filed the above identified patent application which included a copy of Appendix H on Microfiche. The Examiner has indicated that microfiche is not readable by Examiners without special equipment. Applicants submit herewith a copy of Appendix H in paper form.